View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Collection Of Biostatistics Research Archive

Harvard University

Harvard University Biostatistics Working Paper Series

Year 2014 *Paper* 168

Mediation Analysis with Time-Varying Exposures and Mediators

Tyler J. VanderWeele* Eric Tchetgen Tchetgen†

*Harvard University, tvanderw@hsph.harvard.edu

†Harvard School of Public Health

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

http://biostats.bepress.com/harvardbiostat/paper168

Copyright ©2014 by the authors.

Mediation Analysis with Time-Varying Exposures and Mediators

Tyler J. VanderWeele and Eric Tchetgen Tchetgen

Abstract

In this paper we consider mediation analysis when exposures and mediators vary over time. We give non-parametric identification results, discuss parametric implementation, and also provide a weighting approach to direct and indirect effects based on combining the results of two marginal structural models. We also discuss how our results give rise to a causal interpretation of the effect estimates produced from longitudinal structural equation models. When there are no time-varying confounders affected by prior exposure and mediator values, identification of direct and indirect effects is achieved by a longitudinal version of Pearl's mediation formula. When there are time-varying confounders affected by prior exposure and mediator, natural direct and indirect effects are not identified. We define a randomized interventional analogue of natural direct and indirect effects that are identified in this setting. The formula that identifies these effects we refer to as the "mediational g-formula." When there is no mediation, the mediational g-formula reduces to Robins' regular g-formula for longitudinal data. When there are no time-varying confouders affected by prior exposure and mediator values, then the mediational g-formula reduces to a longitudinal version of Pearl's mediation formula. However, the mediational g-formula itself can accomodate both mediation and time-varying confounders.

Mediation Analysis with Time-Varying Exposures and Mediators

Tyler J. VanderWeele and Eric J. Tchetgen Tchetgen

1. Introduction

There has recently been considerable methodologic development on approaches to mediation and pathway analysis from within the causal inference literature (Robins and Greenland, 1992; Pearl, 2001; van der Laan and Petersen, 2008; VanderWeele and Vansteelandt, 2009, 2010; Imai et al., 2010; Valeri and VanderWeele, 2012; Tchetgen Tchetgen and Shpitser, 2012; Lange et al., 2012; Vansteelandt et al., 2012). This work has extended traditional approaches to mediation to settings with interactions and non-linearities and has clarified the no-unmeasured confounding assumptions that suffice for a causal interpretation of direct and indirect effects. Almost all of this literature has considered a single exposure at one point in time, a single mediator, and a single outcome. Often longitudinal data are available and the exposure and the mediator vary over time. There is currently very little work in the causal inference literature with exposures and mediators that very over time. Only a few papers in the causal inference briefly touch on such settings with longitudinal data (van der Laan and Petersen, 2008; VanderWeele, 2009) and an approach that fully accommodates time-varying exposures and mediators and time-varying confounding is yet to be developed. Although some work has been done in psychology on mediation analysis with longitudinal data (cf. MacKinnon, 2008), this does not fall within a formal causal framework. Some of the difficulty is that the concepts of natural direct and indirect effects (Robins and Greenland, 1992; Pearl, 2001) that have been employed in the causal inference literature on mediation are not identified from the data in many settings involving time-varying exposures and mediators. In particular whenever there is a mediator-outcome confounder affected by the exposure, these natural direct and indirect effects are not non-parametrically identified irrespective of whether data is available on this exposure-induced confounder or not (Avin et al., 2005). In the longitudinal settings such exposure-induced confounding may be very common. In this paper we propose an approach to pathway analysis that can be used in settings with time-varying exposures and mediators. To do so, instead of using the natural direct and indirect effects commonly employed in the literature we use a randomized interventional analgoue of natural direct and indirect effects (cf. Didelez et al., 2006; VanderWeele et al., 2014) that can be identified from longitudinal data under weaker assumptions than the natural direct and indirect effects.

2. Natural Direct and Indirect Effects Versus Randomized Interventional Analgoues

In this section we will review the definitions and identification assumptions for the natural direct and indirect effects defined in the causal inference literature on mediation. We will

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

moreover contrast this to randomized interventional analgoues of natural direct and indirect effects which can be identified under weaker assumptions and which will, in the following section, be extended to settings with time-varying exposures and mediators.

Let A denote the exposure of interest; Y, the outcome and M, the potential mediator, and V a set of baseline covariates not affected by the exposure. For now we will assume that the exposure and mediator only occur at one point in time. We will let Y_a and M_a denote, respectively, the values of the outcome and mediator that would have been observed had exposure A been set to level a. We will let Y_{am} denote the value of the outcome that would have been observed had exposure A been set to level a, and mediator M been set to level m. These counterfactual or potential outcome variables, Y_a, M_a and Y_{am} all presuppose that at least hypothetical interventions on A and M are conceivable. A further assumption is often generally made, sometimes referred to as the "consistency assumption", that when $A = a$, the counterfactual outcomes Y_a and M_a are, respectively, equal to the observed outcomes Y and M, and likewise when $A = a$ and $M = m$, the counterfactual outcome Y_{am} is equal to Y.

Using these counterfactuals, Robins and Greenland (1992) and Pearl (2001) defined what have since come to be called controlled direct effects and natural direct and indirect effects. The average controlled direct effect, conditional on covariates $V = v$, comparing exposure level $A = a$ with $A = a^*$ and fixing the mediator to level m, is defined by $E[Y_{am} - Y_{a^*m}|v]$ and captures the effect of exposure A on outcome Y, intervening to fix M to m; it may be different for different levels of m. The natural direct effect, conditional on covariates $V = v$, is defined as $E[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|v]$ and differs from controlled direct effects in that the intermediate M is set to the level M_{a^*}, the level that it would have naturally been if the exposure has taken value $A = a^*$. Similarly, the average natural indirect effect, conditional on $V = v$, can be defined as $E[Y_{aM_a} - Y_{aM_{a^*}}|v]$, which compares the effect of the mediator at levels M_a and M_{a^*} on the outcome when exposure is set to $A = a$. Natural direct and indirect effects have the property that a total effect, $E[Y_1 - Y_0|v]$, decomposes into a natural direct and indirect effect: $E[Y_a - Y_{a^*}|v] = E[Y_{aM_a} - Y_{a^*M_{a^*}}|v] = E[Y_{aM_a} - Y_{aM_{a^*}}|v] + E[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|v]$; the decomposition holds even when there are interactions and non-linearities.

In general, stronger no-unmeasured-confounding assumptions are required to identify direct and indirect effects than total effects. On a causal diagram interpreted as a set of non-parametric structural equations (Pearl, 2009), the following four assumptions suffice to identify natural direct and indirect effects from data (Pearl, 2001; Shpitser and VanderWeele, 2011): (i) the effect the exposure A on the outcome Y is unconfounded conditional on V; (ii) the effect the mediator M on the outcome Y is unconfounded conditional on V; (iii) the effect the exposure A on the mediator M is unconfounded conditional on V; and (iv) there is no effect of the exposure that itself confounds the mediator-outcome relationship. The assumptions would hold if the diagram in Figure 1 were a nonparametric structural equation model.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/harvardbiostat/paper168

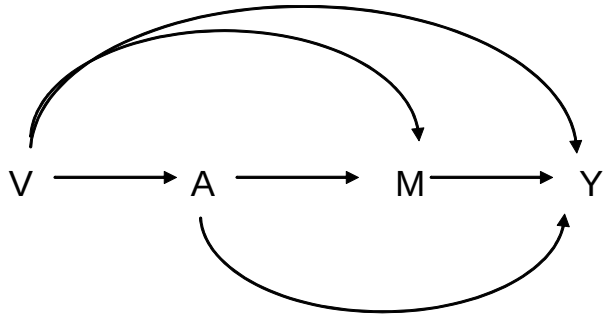


Figure 1. Simple model for mediation.

Only assumptions (i) and (ii) are required to estimate controlled direct effects. Assumptions (i)-(iv) in the text, stated formally in terms of counterfactual independence, are: (i) $Y_{am} \perp\!\!\!\perp A|V$, (ii) $Y_{am} \perp\!\!\!\perp M|\{A, V\}$, (iii) $M_a \perp\!\!\!\perp A|V$, (iv) $Y_{am} \perp\!\!\!\perp M_{a^*}|V$. Under these assumptions natural direct and indirect effects are identified (Pearl, 2001) and given by the following expressions:

$$\begin{aligned} E[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|v] &= \sum\nolimits_m \{E[Y|a, m, v] - E[Y|a^*, m, v]\}P(m|a^*, v). \\ E[Y_{aM_a} - Y_{aM_{a^*}}|v] &= \sum\nolimits_m E[Y|a, m, v]\{P(m|a, v) - P(m|a^*, v)\}. \end{aligned}$$

Importantly, however, note that if there is a mediator-outcome confounder L affected by exposure then assumption (iv) will fail and natural direct and indirect effects will not be identified from the data. Assumption (iv) would thus be violated in Figure 2.

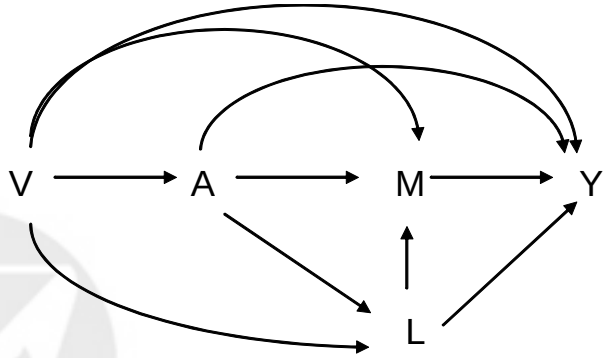


Figure 2. Mediation with a mediator-outcome confounder L that is affected by exposure.

The counterfactual independence assumption (iv) that $Y_{am} \perp\!\!\!\perp M_{a^*}|V$ is also somewhat controversial for other reasons. Although it will hold in the causal diagram in Figure 1 if this diagram is interpreted as a non-parametric structural equation model as in Pearl (2009), there are other interpretations of causal diagrams wherein assumption (iv) may fail even in Figure 1 (Robins, 2003; Robins and Richardson, 2010).

Even if this assumption, that $Y_{am} \perp\!\!\!\perp M_{a^*}|V$, fails, an analogue of natural direct and indirect effects, based on randomized interventions, can be identified from the data under

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

assumptions (i)-(iii) alone. We will conclude this section with a discussion of these randomized interventional analogues of natural direct and indirect effects and in the following section we will consider longitudinal extensions of these effects. These randomized interventional analogues are essentially equivalent to those proposed by Didelez et al. (2006) and Geneletti (2007).

Let $G_{a|v}$ denote a random draw from the distribution of the mediator amongst those with exposure status a conditional on $V = v$. The effect $E(Y_{aG_{a|v}}) - E(Y_{aG_{a^*|v}})$ is then the effect on the outcome of randomly assigning an individual who is given the exposure to a value of the mediator from the distribution of the mediator amongst those given exposure versus not given exposure (conditional on the covariates); this is an effect through the mediator. Next consider the effect $E(Y_{aG_{a^*|v}}) - E(Y_{a^*G_{a^*|v}})$; this is a direct effect comparing exposure versus no exposure with the mediator in both cases randomly drawn from the distribution of the population when given no exposure (conditional on the covariates). Finally, the effect $E(Y_{aG_{a|v}}) - E(Y_{a^*G_{a^*|v}})$ compares the expected outcome when having the exposure with the mediator randomly drawn from the distribution of the population when given the exposure (conditional on covariates) to the expected outcome when not having the exposure with the mediator randomly drawn from the distribution of the population when not exposed. With effects thus defined we have the decomposition: $E(Y_{aG_{a|v}}) - E(Y_{a^*G_{a^*|v}}) = \{E(Y_{aG_{a|v}}) - E(Y_{aG_{a^*|v}})\} + \{E(Y_{aG_{a^*|v}}) - E(Y_{a^*G_{a^*|v}})\}$ so that the overall effect decomposes into the sum of the effect through the mediator and the direct effect. These are not the natural direct and indirect effects considered earlier but are instead analogues arising from fixing the mediator for each individual, not to the level it would have been that for individual under a particular exposure, but rather, to a level that is randomly chosen from the distribution of the mediator amongst all of those with a particular exposure. These effects are identified under assumptions (i)-(iii) alone (VanderWeele et al., 2014). Under these assumptions (i)-(iii) the randomized interventional analogues, $\{E(Y_{aG_{a^*|v}}) - E(Y_{a^*G_{a^*|v}})\}$ and $\{E(Y_{aG_{a|v}}) - E(Y_{aG_{a^*|v}})\}$, are in fact identified by the same empirical expression as those given above for natural direct and indirect effects. Note that assumption (iv) is not necessary for the identification of these randomized interventional effects; it is not necessary because the mediator is being fixed to a level that is randomly chosen from the distribution of the mediator amongst all of those with a particular exposure, rather than fixed to the level it would have been for that individual under a different exposure status. Because assumptions (iv) is not necessary these randomized interventional analogues of natural direct and indirect effects are also identified in interpretation of causal diagrams (Robins and Richardson, 2010) other than Pearl's non-parametric structural equations (cf. VanderWeele et al., 2014). Moreover, even if there is a mediator-outcome confounder affected by the exposure as in Figure 2, the randomized interventional analogues may still be identified from the data but the empirical expressions equal to these effects no longer coincide with that given above for natural direct and indirect effects. They are instead, if Figure 2 is a causal diagram, given by (VanderWeele, et al., 2014):

$$E(Y_{aG_{a^*|v}}) - E(Y_{a^*G_{a^*|v}}) = \sum_{l,m} \{E[Y|a,l,m,v]P(l|a,v) - E[Y|a^*,l,m,v]P(l|a^*,v)\}P(m|a^*,v)$$
$$E(Y_{aG_{a|v}}) - E(Y_{aG_{a^*|v}}) = \sum_{l,m} E[Y|a,l,m,v]P(l|a,v)\{P(m|a,v) - P(m|a^*,v)\}.$$

http://biostats.bepress.com/harvardbiostat/paper168

3. Time-Varying Exposures and Mediators and the Mediational G-Formula

Suppose now that the exposure, mediators and possibly confounding variables vary over time. Let $(A(1), ..., A(T))$, $(M(1), ..., M(T))$, and $(L(1), ..., L(T))$ denote values of the exposures, mediator, and time-varying confounders at periods $0, ..., T$, with initial baseline covariates V, and subsequent temporal ordering $A(t)$, $M(t)$, $L(t)$. We will revisit this question of temporal ordering again later in the paper. The relationships among the variables are given in Figure 3.

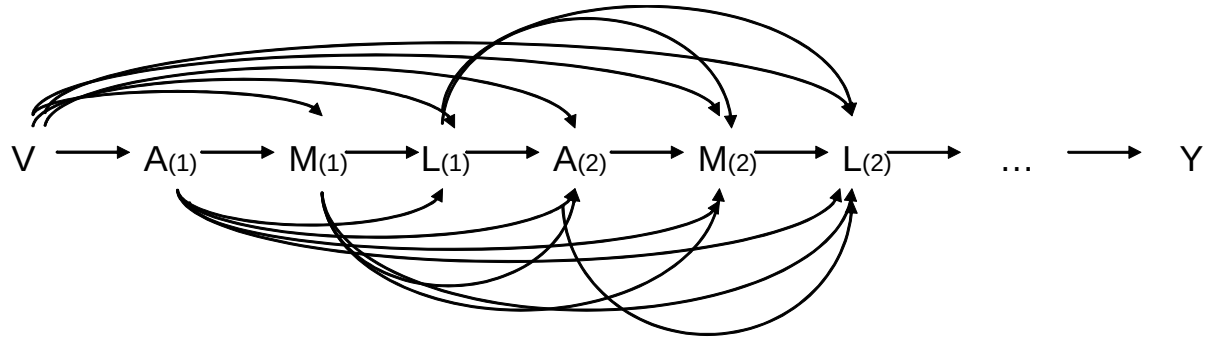


Figure 3. Time-varying mediation with ordering of variables of A(t), M(t), L(t).

For any variable W, let $\overline{W}(t) = (W(1), ..., W(t))$ and let $\overline{W} = \overline{W}(T) = (W(1), ..., W(T))$. Let $\underline{W}(t) = (W(t), ..., W(T)$. By convention, we let $W(t)$ denote the empty set for $t \leq 0$. Let $Y_{\overline{am}}$ be the counterfactual outcome if $\overline{A}$ were set to $\overline{a}$ and if $\overline{M}$ were set to $\overline{m}$. Let $M_{\overline{a}}(t)$ be the counterfactual value of $M(t)$ if $\overline{A}$ were set to $\overline{a}$. We assume consistency that when $\overline{A} = \overline{a}$ we have $M_{\overline{a}}(t) = M(t)$ and $Y_{\overline{a}}(t) = Y(t)$ and when $\overline{A} = \overline{a}$ and $\overline{M} = \overline{m}$ we have $Y_{\overline{am}} = Y$.

Note that if the entire vector $A = (A(1), ..., A(T))$ is taken as the exposure and $M = (M(1), ..., M(T))$ is taken as the mediator then the variable $L(1)$ is itself affected by the exposure (namely, by $A(1)$) and in turn confounds the mediator-outcome relationship between $M(2)$ and Y. From this it follows that natural direct and indirect effects are not identified in this setting (Avin et al., 2005). However, identification of randomized interventional analogues may once again be possible.

Let $\overline{G}_{\overline{a}|v}(t)$ denote a random draw from the distribution of the mediator $\overline{M}(t)$ that would have been observed in the population with baseline covariates $V = v$ if exposure status $\overline{A}$ had been fixed to $\overline{a}$. Let $\overline{a}$ and $\overline{a}^*$ be two distinct exposure histories. We once again have a decomposition, even with time-varying exposures and mediators: $E(Y_{\overline{a}\overline{G}_{\overline{a}|v}(t)}|v) - E(Y_{\overline{a}^* G_{\overline{a}^*|v}}|v) = \{E(Y_{\overline{a} G_{\overline{a}|v}}|v) - E(Y_{\overline{a} G_{\overline{a}^*|v}}|v)\} + \{E(Y_{\overline{a} G_{\overline{a}^*|v}}|v) - E(Y_{\overline{a}^* G_{\overline{a}^*|v}}|v)\}$.

Although these randomized interventional analogues defined here are not identical with natural direct and indirect, they are in some sense the best we may be able to do as the natural direct and indirect effects themselves will not be identified when a mediator-outcome confounder is affected by the exposure; in such settings the randomized interventional analogues are then all that we can estimate. Moreover, several further comments merit attention. First, these randomized interventional analogues do in some sense capture mediated

Hosted by The Berkeley Electronic Press

effects and pathways; the randomized interventional analogues of the natural indirect effect, $\{E(Y_{\overline{a}G_{\overline{a}|v}}|v) - E(Y_{\overline{a}G_{\overline{a}^*|v}}|v)\}$, will be non-zero only if the exposure changes the distribution of the mediator and that change in the distribution of the mediator changes the outcome. Second, when there are no mediator-outcome confounders affected by the exposure, it will be seen below that the randomized interventional analogues in fact do coincide with natural direct and indirect effects; thus when the latter effects are identified the randomized interventional analogues in fact capture these effects. Third, when natural direct and indirect effects are not identified, it will only be in extremely pathological settings that the randomized analogue is non-zero, but there are in fact no natural indirect effects. For that to occur, it would be necessary that the exposure affects the mediator for a completely different set of individual than for whom the mediator affects the outcome i.e. there is no overlap in those for whom the exposure affects the mediator and for whom the mediator affects the outcome. Conversely for there to be a non-zero natural indirect effect with a zero randomized interventional analogue of that effect would essentially require exact cancellations to occur.

Finally, there are arguably some settings in which the randomized interventional analgoues are in fact what is of principal substantive interest, rather than the natural direct and indirect effect. Suppose we were interested in whether a racial health disparity (race constituting the exposure, and health the outcome) was mediated by differences in socioeconomic distributions. The natural direct and indirect effects would entail hypothetical interventions on the mediator of fixing a black individual's socioeconomic status to what it would have been had they been white. Counterfactual queries of the form of what a black individual's socioeconomic status would have been had they been of a different race strike most people as strange or meaningless. However, the randomized interventional analogues arguably involve much less problematic comparisons. The randomized interventional analogue of the natural direct effect say, essentially entails just asking how much of a racial health disparity would remain if we fixed the socioeconomic distributions of the black individual to be the same distribution as that of the white individuals. By randomly fixing the distributions to equal one another, we avoid peculiar counterfactuals of the form of what would have happened to an individual had they been of a different race. See VanderWeele and Robinson (2014) for further discussion. Thus, in some cases at least, the randomized interventional analogues are not simply a second-best alternative to natural direct and indirect effects, but are themselves arguably the causal effects of interest.

Suppose now that at each time, conditional on the past, the exposure-outcome-, mediator-outcome-, and exposure-mediator- relationships are unconfounded. Formally, analogous to (i)-(iii): for all t, (i†) $Y_{\overline{am}} \perp\!\!\!\perp A(t)|\overline{A}(t-1), \overline{M}(t-1), \overline{L}(t-1), V$ and (ii†) $Y_{\overline{am}} \perp\!\!\!\perp M(t)|\overline{A}(t), \overline{M}(t-1), \overline{L}(t-1), V$ and (iii†) $M_{\overline{a}}(t) \perp\!\!\!\perp A(t)|\overline{A}(t-1), \overline{M}(t-1), \overline{L}(t-1), V$. It can be shown that although natural direct and indirect effects are not in general identified in this setting, the randomized interventional analogues, $\{E(Y_{\overline{a}G_{\overline{a}|v}}|v) - E(Y_{\overline{a}G_{\overline{a}^*|v}}|v)\}$ and

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/harvardbiostat/paper168

$\{E(Y_{\overline{a}G_{\overline{a}^*|v}}|v) - E(Y_{\overline{a}^*G_{\overline{a}^*|v}}|v)\}$, are identified. This is because:

$$\begin{aligned} & E[Y_{\overline{a}G_{\overline{a}^*|v}}|v] \\ = & \sum\nolimits_{\overline{m}} E[Y_{\overline{am}}|G_{\overline{a}^*|v} = \overline{m}, v]P(G_{\overline{a}^*|v} = \overline{m}|v) \\ = & \sum\nolimits_{\overline{m}} E[Y_{\overline{am}}|v]P(M_{\overline{a}^*} = \overline{m}|v) \end{aligned}$$

and applying the g-formula (Robins, 1986) to each of $E[Y_{\overline{am}}|v]$ and $P(M_{\overline{a}^*} = \overline{m}|v)$ we obtain

$$\begin{aligned} & \sum\nolimits_{\overline{m}} \sum\nolimits_{\overline{l}(T-1)} E[Y|\overline{a}, \overline{m}, \overline{l}, v] \prod\nolimits_{t=1}^{T-1} P\{\overline{l}(t)|\overline{a}(t), \overline{m}(t), \overline{l}(t-1), v)\} \\ & \times \sum\nolimits_{\overline{l}^\dagger(T-1)} \prod\nolimits_{t=1}^{T} P\{M(t)|\overline{a}^*(t), \overline{m}(t-1), \overline{l}^\dagger(t-1), v\} P\{\overline{l}^\dagger(t-1)|\overline{a}^*(t-1), \overline{m}(t-1), \overline{l}^\dagger(t-2), v)\}. \end{aligned} \tag{1}$$

An alternative derivation is also given in the appendix.

We refer to this final expression in (1) as the mediational g-formula. We will denote this quantity by $Q(\overline{a}, \overline{a}^*)$. Our randomized interventional analogues of natural direct and indirect effects are under assumptions (i†)-(iii†) then given by

$$\begin{aligned} E(Y_{\overline{a}G_{\overline{a}|v}}|v) - E(Y_{\overline{a}G_{\overline{a}^*|v}}|v) &= Q(\overline{a}, \overline{a}) - Q(\overline{a}, \overline{a}^*) \\ E(Y_{\overline{a}G_{\overline{a}^*|v}}|v) - E(Y_{\overline{a}^*G_{\overline{a}^*|v}}|v) &= Q(\overline{a}, \overline{a}^*) - Q(\overline{a}^*, \overline{a}^*) \end{aligned}$$

Note that if $\overline{L}$ is empty as in Figure 4 then the mediational g-formula reduces to

$$Q(\overline{a}, \overline{a}^*) = \sum\nolimits_{\overline{m}} E[Y|\overline{a}, \overline{m}, v] \prod\nolimits_{t=1}^{T} P\{M(t)|\overline{a}^*(t), \overline{m}(t-1), v\}.$$

We show in the appendix that if $\overline{L}$ is empty then, under a non-parametric structural equation model, natural direct effects are identified by the mediational g-formula and are equal to $Q(\overline{a}, \overline{a}^*) - Q(\overline{a}^*, \overline{a}^*)$ and natural indirect effects are identified by the mediational g-formula and are equal to $Q(\overline{a}, \overline{a}) - Q(\overline{a}, \overline{a}^*)$.

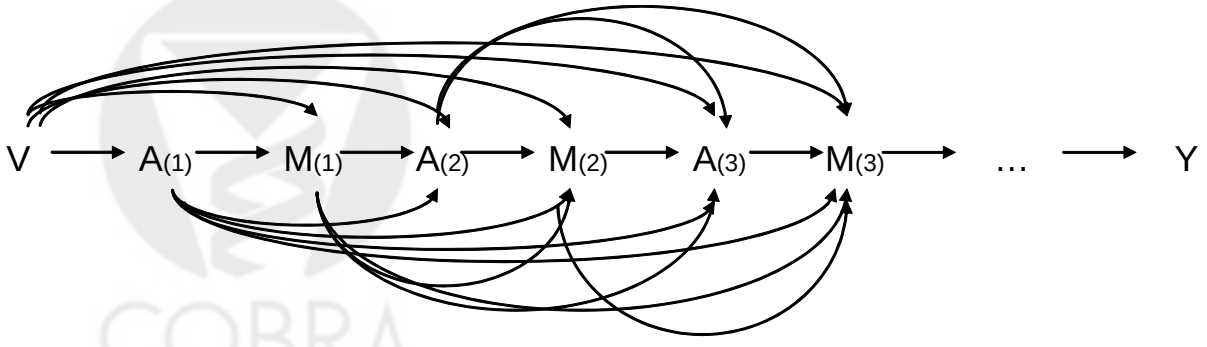


Figure 4. Time-varying exposures and mediators, with no time-varying confounders.

In other words if $\overline{L}$ is empty then the empirical expressions that suffice to identify the randomized interventional analogues of natural direct and indirect effects under assumptions

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

(i)-(iii) in fact also in this setting identify the natural direct and indirect effects as well by a time-varying analogue of Pearl's "mediation formula" (Pearl, 2012). However, even when $\overline{L}$ is not empty so we cannot identify the natural direct and indirect effects themselves, we still can, under assumptions (i†)-(iii†) identify the randomized interventional analogues of the natural direct and indirect effects.

Note also that if M were empty then the expression in (1) simply reduces to:

$$\sum_{\overline{m}} \sum_{\overline{l}(T-1)} E[Y|\overline{a}, \overline{l}, v] \prod_{t=1}^{T-1} P\{\overline{l}(t)|\overline{a}(t), \overline{l}(t-1), v)\}$$

because, with M empty, $\sum_{\overline{l}^\dagger(T-1)} \prod_{t=1}^{T} P\{\overline{l}^\dagger(t-1)|\overline{a}^*(t-1), \overline{l}^\dagger(t-2), v)\} = 1$. Thus with M empty, the formula in (1) simply reduces to the regular g-formula of Robins (1986). We see then that, on the one hand, if there is no-time-varying confounding the "mediational g-formula" in (1) reduces to the time-varying analogue of the mediational formula. And if, on the other hand, there is no mediation, then the "mediational g-formula" reduces to the regular g-formula.

We now consider some variations on this approach. First, suppose instead that after the initial baseline covariates V, the subsequent temporal ordering of the variables were $A(t)$, $L(t)$, $M(t)$, as in Figure 5, and that analogous to (i†)-(iii†) we have that: for all t, (i‡) $Y_{\overline{am}} \perp\!\!\!\perp A(t)|\overline{A}(t-1), \overline{M}(t-1), \overline{L}(t-1), V$ and (ii‡) $Y_{\overline{am}} \perp\!\!\!\perp M(t)|\overline{A}(t), \overline{M}(t-1), \overline{L}(t), V$ and (iii‡) $M_{\overline{a}}(t) \perp\!\!\!\perp A(t)|\overline{A}(t-1), \overline{M}(t-1), \overline{L}(t-1), V$.

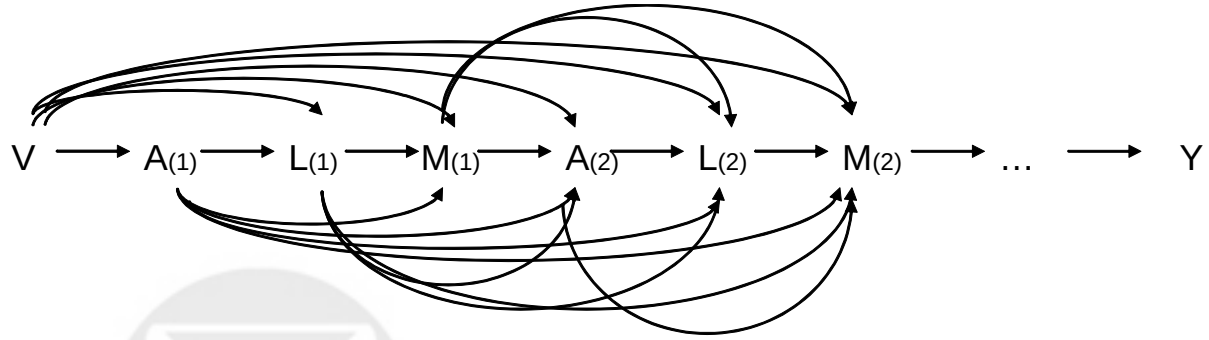


Figure 5. Time-varying mediation with variable ordering A(t), L(t), M(t).

Under assumptions (i‡)-(iii‡) we would then have:

$$\begin{aligned}
& E[Y_{\overline{a}G_{\overline{a}^*|v}}|v] \\
= & \sum_{\overline{m}} E[Y_{\overline{am}}|G_{\overline{a}^*|v} = \overline{m}, v] P(G_{\overline{a}^*|v} = \overline{m}|v) \\
= & \sum_{\overline{m}} E[Y_{\overline{am}}|v] P(M_{\overline{a}^*} = \overline{m}|v) \\
= & \sum_{\overline{m}} \sum_{\overline{l}(T-1)} E[Y|\overline{a}, \overline{m}, \overline{l}, v] \prod_{t=1}^{T-1} P\{\overline{l}(t)|\overline{a}(t), \overline{m}(t-1), \overline{l}(t-1), v)\} \\
& \times \sum_{\overline{l}^\dagger(T-1)} \prod_{t=1}^{T} P\{M(t)|\overline{a}^*(t), \overline{m}(t-1), \overline{l}^\dagger(t), v\} P\{\overline{l}^\dagger(t)|\overline{a}^*(t), \overline{m}(t-1), \overline{l}^\dagger(t-1), v)\}
\end{aligned}$$

COBRA
BEPRESS
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/harvardbiostat/paper168

where the final equality again follows by applying the g-formula of Robins (1986).

As another variation instead of considering randomized interventions that fix the mediator $\overline{M}$ for each individual to a value randomly drawn the distribution in the subpopulation with baseline covariates $V = v$ if $\overline{A}$ had been fixed to $\overline{a}^*$, we could instead consider randomizing the mediator $\overline{M}$ for each individual to the value randomly drawn the distribution in the entire population if $\overline{A}$ had been fixed to $\overline{a}^*$. We then let $\overline{G}_{\overline{a}}(t)$ denote a random draw from the distribution of the mediator $\overline{M}(t)$ that would have been observed in the population if exposure $\overline{A}$ had been fixed to $\overline{a}$ and we have the decomposition: $E(Y_{\overline{a}\overline{G}_{\overline{a}}(t)}) - E(Y_{\overline{a}^* G_{\overline{a}^*}}) = \{E(Y_{\overline{a}G_{\overline{a}}}) - E(Y_{\overline{a}G_{\overline{a}^*}})\} + \{E(Y_{\overline{a}G_{\overline{a}^*}}|v) - E(Y_{\overline{a}^* G_{\overline{a}^*}})\}$. Using under assumptions (i†)-(iii†) we have: $E[Y_{\overline{a}G_{\overline{a}^*}}] =$

$$\sum_{\overline{m}} \sum_{\overline{l}(T-1)} E[Y|\overline{a}, \overline{m}, \overline{l}, v] \prod_{t=1}^{T-1} P\{\overline{l}(t)|\overline{a}(t), \overline{m}(t), \overline{l}(t-1), v)\} P(v)$$
$$\times \sum_{\overline{l}^\dagger(T-1)} \prod_{t=1}^{T} P\{M(t)|\overline{a}^*(t), \overline{m}(t-1), \overline{l}^\dagger(t-1), v\} P\{\overline{l}^\dagger(t-1)|\overline{a}^*(t-1), \overline{m}(t-1), \overline{l}^\dagger(t-2), v)\} P(v)$$

and under assumptions assumptions (i‡)-(iii‡) we would then have: $E[Y_{\overline{a}G_{\overline{a}^*}}]$

$$\begin{aligned} = \quad & \sum_{\overline{m}} \sum_{\overline{l}(T-1)} E[Y|\overline{a}, \overline{m}, \overline{l}, v] \prod_{t=1}^{T-1} P\{\overline{l}(t)|\overline{a}(t), \overline{m}(t-1), \overline{l}(t-1), v)\} P(v) \\ & \times \sum_{\overline{l}^\dagger(T-1)} \prod_{t=1}^{T} P\{M(t)|\overline{a}^*(t), \overline{m}(t-1), \overline{l}^\dagger(t), v\} P\{\overline{l}^\dagger(t)|\overline{a}^*(t), \overline{m}(t-1), \overline{l}^\dagger(t-1), v)\} P(v). \end{aligned}$$

Note that in all of the above variations, we have fixed the entire mediator $\overline{M}$ to a random draw from the mediator vector under a particular exposure history. As yet another alternative, though one we argue is not suitable for mediation analysis, we could have at each time t, fixed the mediator $M(t)$ for that time t, to a random draw from the mediator distribution under a particular exposure history up to that point in time t. Said another way, we could have fixed to mediator to a random draw from the mediator distribution under a specific exposure distribution marginally, rather than jointly as in all the variations considered above. If we had proceeded in this manner the identifying expression would have differed. Doing so, however, we argue does not adequately allow for the analysis of pathways. To see this, consider the following example: suppose we were interested in assessing the extent to which the effect of marital status (which may be time-varying) on income is mediated by time-varying health status. Suppose that different individuals with different marital status histories have different health trajectories, and that at least some individuals have consistently poor health over time if and only if in the unmarried state, but that the vast majority are healthy over time in either marital state. Suppose that it is only a long-term poor health trajectory that substantially affects income. If we were to randomize the entire mediator vector to a draw from the health trajectory distribution of those who were unmarried then some of these trajectories randomly drawn would be consistently low and would adversely affect income. Using the approaches described above we would see that some of the effect of marital status on income was mediated by preventing the consistently low health trajectories. However, if we were instead to randomize the mediator marginally at each time point to a random draw of the distribution of the unmarried population, the probability of obtaining a health trajectory that was consistently low over time would be very very small (since at each time

COBRA
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

the majority are in the healthy state and thus to get a consistently low health trajectory would require low probability events at each of the individual time points). Consequently, if we were to randomize the mediator marginally at each time point, far fewer individuals in a setting in which the mediator were randomized marginally at each time according to the unmarried distribution would have a health trajectory which was consistently low at all time points than was actually the case with the actual unmarried population and thus there would be few individual for whom income was substantially adversely affected by health and we would for the most part miss those pathways by which marital status affects income through consistently low health trajectories. To assess such pathways we need to randomize the mediator jointly at all time points to a random draw from the distribution of those with a particular exposure history, as in the approaches described above.

4. Estimation Using Marginal Structural Models

One possible estimation approach would be to use the identification formula in (1) and fit parametric models for each of $E[Y|\overline{a},\overline{m},\overline{l},v]$, $P\{\overline{l}(t)|\overline{a}(t),\overline{m}(t),\overline{l}(t-1),v)\}$, and $P\{M(t)|\overline{a}(t),\overline{m}(t-1),\overline{l}(t-1),v\}$. This estimation approach is sometimes called a g-computation approach and is described in the setting of time-varying exposures outside of the context of mediation elsewhere. We will in fact consider one such approach in the context of MacKinnon's three wave longitudinal mediation model (MacKinnon, 2008) in the following section. However, in general such an approach requires fitting many parametric models and it can sometimes be difficult to specify these models so that they are compatiable with one another and compatiable with the null hypothesis of no effect; these problems are discussed in the setting of time-varying exposures outside of the context of mediation elsewhere. In this section we will instead develop a more parsimonious approach to estimating the randomized interventional analogues of natural direct and indirect effects using marginal structural models and inverse probability of treatment weighting (Robins et al., 2000).

One reasonably straightforward approach entails positing a pair of marginal structural models (MSMs) for $E[Y_{\overline{am}}|v]$ and $P(M_{\overline{a}^*}=\overline{m}|v)$, which we shall denote $E[Y_{\overline{am}}|v;\beta_y]$ and $P(M_{\overline{a}^*}=\overline{m}|v;\beta_m)$. These models can in turn be used to evaluate direct and indirect effects using the following expression previously derived:

$$\begin{aligned} & E[Y_{\overline{a}G_{\overline{a}^*|v}}|v;\beta_y,\beta_m] \\ = & \sum\nolimits_{\overline{m}} E[Y_{\overline{am}}|v;\beta_y]P(M_{\overline{a}^*}=\overline{m}|v;\beta_m) \end{aligned}$$

Consider a scenario, in which Y is a continuous outcome, and V is empty. We assume the following simple marginal structural linear regression model for the outcome:

$$E[Y_{\overline{am}};\beta_y]=\beta_{y0}+\beta_{ya}\text{cum}\left(\overline{a}\right)+\beta_{ym}\text{cum}\left(\overline{m}\right) \tag{2}$$

where $\beta_y=\left\{\beta_{y0}\left(t\right),\beta_{ym}\left(t\right),\beta_{ya}\left(t\right)\right\}$, and $\text{cum}\left(\overline{a}\right)=\sum_{t<T}a(t)$ and $\text{cum}\left(\overline{m}\right)=\sum_{t<T}m(t)$ are the cumulative totals of $\overline{A}$ and $\overline{M}$ respectively. This MSM assumes that the joint effects of $\overline{M}$ and $\overline{A}$ is cumulative, with a single parameter β_{ym} encoding the effect of the M

http://biostats.bepress.com/harvardbiostat/paper168

process through $\text{cum}(\overline{m}) = \sum_{t<T} m(t)$ and β_{ya} encoding the effect of the A process through $\text{cum}(\overline{a}) = \sum_{t<T} a(t)$. For continuous M or A, the model essentially states that the joint effects of $\overline{M}$ and $\overline{A}$ on Y operate strictly through their respective historical average levels, and that these two processes do not interact on the additive scale. A more flexible model could also be specified to account for possibly more complex dose-response relationships between $(\overline{a}, \overline{m})$ and $Y_{\overline{am}}$ and interactions between $\overline{m}$ and $\overline{a}$ could also be specified. Together with Model (2), suppose that the following MSM model holds for the mediator process

$$\text{g}^{-1}\left\{E(M_{\overline{a}}(t)); \beta_m\right\} = \beta_{m0}(t) + \beta_{ma}(t)\,\text{avg}\left(\overline{a}(t-1)\right) \tag{3}$$

where $g^{-1}(\cdot)$ is a link function, and $\beta_m = \left\{\beta_{m0}(t), \beta_{ma}(t) : t\right\}$ and $\text{avg}\left(\overline{a}(t-1)\right) = \sum_{t<T} a(t)/T$. It is easy to verify that models (2) and (3) induce the model

$$\begin{aligned} & E[Y_{\overline{a}G_{\overline{a}^*|v}}|v; \beta_y, \beta_m] \\ = & \sum\nolimits_{\overline{m}} E[Y_{\overline{am}}|v; \beta_y] P(M_{\overline{a}^*} = \overline{m}|v; \beta_m) \\ = & \beta_{y0} + \beta_{ym}\left(\sum_{t<T} g\left(\beta_{m0}(t) + \beta_{ma}(t)\,\text{avg}\left(\overline{a}^*(t-1)\right)\right)\right) + \beta_{ya}\text{cum}(\overline{a}) \end{aligned}$$

In the special case where $M(t)$ is continuous, so that g^{-1} may be taken to be the identity link, one obtains the following expression for the direct effect:

$$\left\{E(Y_{\overline{a}G_{\overline{a}^*}}|v) - E(Y_{\overline{a}^*G_{\overline{a}^*}})\right\} = \beta_{ya}\left\{\text{cum}(\overline{a}) - \text{cum}(\overline{a}^*)\right\},$$

and for the indirect effect:

$$\left\{E(Y_{\overline{a}G_{\overline{a}}}) - E(Y_{\overline{a}G_{\overline{a}^*}})\right\} = \sum_{t<T} \beta_{ym}\beta_{ma}(t)\left\{\text{avg}\left(\overline{a}(t-1)\right) - \text{avg}\left(\overline{a}^*(t-1)\right)\right\}$$

Interestingly, the expression in the above display further simplifies when $\beta_{ma}(t) = \beta_{ma}$ is assumed to be constant, $a^*(t-1) = 0$ and $a(t) = 1$ for all t, producing the following simple expression for the indirect effect:

$$\left\{E(Y_{\overline{a}G_{\overline{a}}}) - E(Y_{\overline{a}G_{\overline{a}^*}})\right\} = \beta_{ym}\beta_{ma}$$

For estimation, standard inverse probability weighting may be used to estimate $\left(\beta_y, \beta_m\right)$, however, construction of the weights varies somewhat with the underlying identifying assumptions. Specifically, suppose that assumptions (i†)-(iii†) hold, then a consistent estimate of β_y under model (2) can be obtained by weighted least squares regression of Y on $\left(\text{cum}\left(\overline{M}\right), \text{cum}\left(\overline{A}\right)\right)$ with estimated weight equal to

$$\prod\nolimits_{t=1}^{T-1} \widehat{P}\{A(t), M(t)|\overline{A}(t-1), \overline{M}(t-1), \overline{L}(t-1), V)\}^{-1}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

where

$$\begin{aligned} & \widehat{P}\{A(t), M(t)|\overline{A}(t-1), \overline{M}(t-1), \overline{L}(t-1), V)\} \\ = \; & \widehat{P}\{M(t)|\overline{A}(t), \overline{M}(t-1), \overline{L}(t-1), V)\}\widehat{P}\{A(t)|\overline{A}(t-1), \overline{M}(t-1), \overline{L}(t-1), V)\} \end{aligned}$$

is a maximum likelihood estimate of $P\{A(t), M(t)|\overline{A}(t-1), \overline{M}(t-1), \overline{L}(t-1), V)\}$ under a standard parametric model. The parameter $\beta_m(t)$ of the second MSM (3) is likewise estimated via inverse probability weighted regression with weight

$$\prod_{s=1}^{t} \widehat{P}\{A(t)|\overline{A}(t-1), \overline{M}(t-1), \overline{L}(t-1), V)\}^{-1}$$

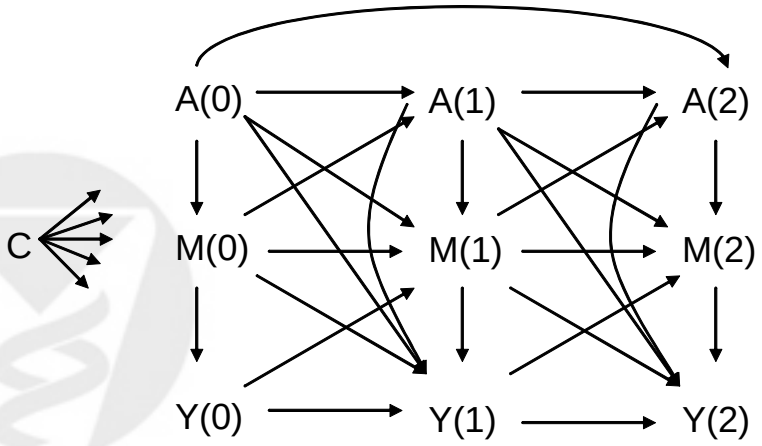
It is straightforward to modify the weights for estimation under the alternative identifying assumptions ($i^{\ddagger}$)-($iii^{\ddagger}$). Specifically, estimation of β_y under model (2) would instead use the following set of weights

$$\left[\prod_{t=1}^{T-1} \widehat{P}\{M(t)|\overline{A}(t), \overline{M}(t-1), \overline{L}(t), V)\}\widehat{P}\{A(t)|\overline{A}(t-1), \overline{M}(t-1), \overline{L}(t-1), V)\}\right]^{-1}$$

while estimation of $\beta_m(t)$ in the second MSM (3) would use the same set of weights as above. In either situation inference can proceed using the nonparametric bootstrap, to appropriately account for variation due to estimation of the weights.

5. A Counterfactual Analysis of MacKinnon's Three-Wave Mediation Model

MacKinnon (2008) considered a three-wave mediation model with linear structural equations as depicted in Figure 6.



We relabel indices somewhat to correspond to the notation of this chapter, and also add a set of baseline covariates C, but otherwise the model considered here is MacKinnon's model (MacKinnon, 2008, pp. 204-206, Autoregressive Model III). We let $A(0)$, $M(0)$ and $Y(0)$ denote baseline values of A, M and Y that could be included in the baseline covariates C

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/harvardbiostat/paper168

but are given here to make clearer the relation with MacKinnon (2008). Consider then the following regression models:

$$\begin{aligned}
E[M(1)|m(0), y(0), \overline{a}(1), c] &= \beta_{10} + \beta_{11}a(0) + \beta_{12}a(1) + \beta_{13}m(0) + \beta_{14}y(0) + \beta'_{15}c \\
E[M(2)|\overline{m}(1), \overline{y}(1), \overline{a}(2), c] &= \beta_{20} + \beta_{21}a(1) + \beta_{22}a(2) + \beta_{23}m(1) + \beta_{24}y(1) + \beta'_{25}c \\
E[Y(1)|\overline{m}(1), y(0), \overline{a}(1), c] &= \theta_{10} + \theta_{11}a(0) + \theta_{12}a(1) + \theta_{13}m(0) + \theta_{14}m(1) + \theta_{15}y(0) + \theta'_{16}c \\
E[Y(2)|\overline{m}(2), \overline{y}(1), \overline{a}(2), c] &= \theta_{20} + \theta_{21}a(1) + \theta_{22}a(2) + \theta_{23}m(1) + \theta_{24}m(2) + \theta_{25}y(1) + \theta'_{26}c.
\end{aligned}$$

Note that in these models, the mediator and the outcome depend only on the two most recent past exposure values. The mediator model depends only on the most recent past mediator value and the most recent past outcome value. The outcome model depends on the two most recent mediator values and the most recent outcome value.

We show that under assumptions (i†)-(iii†) with $V = (C, A(0), M(0), Y(0))$ and $L(1) = Y(1)$, with two intervention periods, $A(1)$ and $A(2)$, the randomized interventional analogues of the natural direct and indirect effects are given by:

$$\begin{aligned}
E(Y_{\overline{a}G_{\overline{a}^*|v}}|v) - E(Y_{\overline{a}^*G_{\overline{a}^*|v}}|v) &= (\theta_{21} + \theta_{12}\theta_{25})[a(1) - a^*(1)] + \theta_{22}[a(2) - a^*(2)] \\
\{E(Y_{\overline{a}G_{\overline{a}|v}}|v) - E(Y_{\overline{a}G_{\overline{a}^*|v}}|v)\} &= \{\theta_{23}\beta_{12} + \theta_{25}\theta_{14}\beta_{12} + \beta_{21}\theta_{24} + \beta_{24}\theta_{12}\theta_{24}\}[a(1) - a^*(1)] \\
&\quad + \beta_{22}\theta_{24}[a(2) - a^*(2)].
\end{aligned}$$

The first expression is the randomized interventional analogue of the natural direct effect with time-varying exposure and mediator and the second expression is the randomized interventional analogue of the natural indirect effect with time-varying exposure and mediator. A proof of this is given in the Appendix.

There is arguably a two-fold advantage of using data like that in Figure 5 and using a modeling approach like that described above, over simply applying the standard methods for mediation to one point in time e.g. using the variables $A(1), M(1), Y(1)$. First, by having multiple waves of data, we can control for baseline levels of the exposure, mediator and outcome, i.e. for $A(0)$, $M(0), Y(0)$. This is potentially important because such baseline values of the exposure, mediator and outcome may serve as the most important confounders for the effects of subsequent values of exposure and mediator on the outcome. By including such baseline values of the exposure, mediator and outcome, in our covariate set, our confounding assumptions required for a causal interpretation of our estimates are rendered much more plausible. Second, by using multiple waves of subsequent exposure and mediator and outcome data (i.e. by using $A(1), M(1), Y(1), A(2), M(2), Y(2)$ rather than just $A(1), M(1), Y(1)$) we may be able to more fully capture the dynamics of mediation over time. For example we can pick up, in our indirect effect estimates, mediated effects of $A(1)$ through $M(1)$ to $Y(2)$ directly and also those from $A(1)$ through $M(1)$ to $Y(1)$ to $Y(2)$ or from $A(1)$ to $M(2)$ to $Y(2)$, etc.

Here we have given a counterfactual analysis of one specific mediational model with three waves of data on the exposure, mediator and outcome (MacKinnon, 2008). A similar approach could in principle be used for other complex longitudinal models often used in the

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

social sciences to provide counterfactual-based interpretations of direct and indirect effect estimates.

6. Discussion

In this paper we have considered methods for time-varying exposures and mediators. One of the challenges here was mediator-outcome confounder affected by the exposure. This can lead to lack of non-parametric identification of longitudinal analogues of natural direct and indirect effects. However we were able to show in this paper that it is still possible estimate randomized interventional analogues of natural direct and indirect effects and these can in fact be used for effect decomposition. These randomized interventional analogues do reduce to the natural direct and indirect effects where there is no mediator-outcome confounder affected by exposure (e.g. when there are no time-varying confounders) but the randomized interventional analogues can be estimated in a broader range of settings even when natural direct and indirect effects are not identified with the data. The methods in this paper thereby extend those in previous chapters to settings with longitudinal data and exposures and mediators that vary over time. Such rich longitudinal data can potentially increase power in the analysis of direct and mediated effects and help better ensure that questions of temporality in thinking about causal effects are clearer.

References

Avin, C., Shpitser, I., and Pearl, J. (2005). Identifiability of path-specific effects. In Proceedings of the International Joint Conferences on Artificial Intelligence, 357-363.

Geneletti, S. (2007). Identifying direct and indirect effects in a non-counterfactual framework. *Journal of the Royal Statistical Soceity, Series B*, 69:199-216.

Didelez, V., Dawid, A. P. and Geneletti, S. (2006) Direct and indirect effects of sequential treatments. In Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence.

Imai, K., Keele, L., Tingley, D. (2010). A general approach to causal mediation analysis. Psychological Methods, 15:309-334.

Lange, T. and Hansen, J.V. (2012). Direct and indirect effects in a survival context. *Epidemiology*, 2011;22:575-581.

MacKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis*. New York: Erlbaum.

Pearl, J. (2001). Direct and indirect effects. In Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence. San Francisco: Morgan Kaufmann, 411-420.

Pearl, J. (2009). Causality: Models, Reasoning, and Inference. Cambridge: Cambridge University Press, 2nd edition.

http://biostats.bepress.com/harvardbiostat/paper168

Pearl, J. (2012). The causal mediation formula - a guide to the assessment of pathways and mechanisms. *Prevention Science*, 13:426-436.

Robins, J.M. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In Highly Structured Stochastic Systems, Eds. P. Green, N.L. Hjort, and S. Richardson, 70-81. Oxford University Press, New York.

Robins, J.M. (1986). A new approach to causal inference in mortality studies with sustained exposure period - application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7:1393-1512.

Robins, J.M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. Epidemiology, 3:143-155.

Robins, J.M and Richardson, T.S. (2010). Alternative graphical causal models and the identification of direct effects. In P. Shrout (Ed.): Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures. Oxford University Press.

Shpitser, I., and Pearl, J. (2008). Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research*, 9:1941-1979.

Shpitser, I. and VanderWeele, T.J. (2011). A complete graphical criterion for the adjustment formula in mediation analysis. *International Journal of Biostatistics*, 7, Article 16:1-24.

Tchetgen Tchetgen, E.J. and Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. Annals of Statistics, 40(3):1816-1845.

Valeri, L. and VanderWeele, T.J. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychological Methods, in press.

van der Laan, M.J. and Petersen, M.L. (2008). Direct effect models. International Journal of Biostatistics, 4: Article 23.

VanderWeele, T.J. (2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*, 20:18-26.

VanderWeele, T.J. and Robinson, W. (2014). On the causal interpretation of race in regressions adjusting for confounding and mediating variables. Epidemiology, in press.

VanderWeele T.J. and Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. Statistics and Its Interface - Special Issue on Mental Health and Social Behavioral Science, 2:457-468.

COBRA
Research Archive

Hosted by The Berkeley Electronic Press

VanderWeele T.J., Vansteelandt, S., and Robins, J.M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology, in press.

VanderWeele, T.J. and Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. American Journal of Epidemiology, 172:1339-1348.

Vansteelandt, S., Bekaert, M. and Lange, T. (2012). Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiologic Methods*, 1:131-158.

Appendix

Alternative derivation of the mediational g-formula

We have that:

$$
\begin{aligned}
& E[Y_{\overline{a}G_{\overline{a}^*|v}}|v] \\
= & \sum_{m(1)} E[Y_{\overline{a}m(1)\underline{G}_{\overline{a}^*|v}(2)}|G_{\overline{a}^*|v}(1)=m(1),v]P\{G_{\overline{a}^*|v}(1)=m(1)|v\} \\
= & \sum_{m(1)} E[Y_{\overline{a}m(1)\underline{G}_{\overline{a}^*|v}(2)}|G_{\overline{a}^*|v}(1)=m(1),a(1),m(1),v]P\{M_{\overline{a}^*}(1)=m(1)|a^*(1),v\} \\
= & \sum_{m(1)} E[Y_{\overline{a}m(1)\underline{G}_{\overline{a}^*|v}(2)}|G_{\overline{a}^*|v}(1)=m(1),a(1),m(1),v]P\{m(1)|a^*(1),v\} \\
= & \sum_{\overline{m}(2)} E[Y_{\overline{a}\overline{m}(2)\underline{G}_{\overline{a}^*|v}(3)}|\overline{G}_{\overline{a}^*|v}(2)=\overline{m}(2),a(1),m(1),v] \\
& \times P\{G_{\overline{a}^*|v}(2)=m(2)|G_{\overline{a}^*|v}(1)=m(1),a(1),m(1),v\}P\{m(1)|a(1),v\} \\
= & \sum_{\overline{m}(2)} E[Y_{\overline{a}\overline{m}(2)\underline{G}_{\overline{a}^*|v}(3)}|\overline{G}_{\overline{a}^*|v}(2)=\overline{m}(2),a(1),m(1),v] \\
& \times P\{G_{\overline{a}^*|v}(2)=m(2)|G_{\overline{a}^*|v}(1)=m(1),a^*(1),m(1),v\}P\{m(1)|a(1),v\} \\
= & \sum_{\overline{m}(2)}\sum_{\overline{l}(1)} E[Y_{\overline{a}\overline{m}(2)\underline{G}_{\overline{a}^*|v}(3)}|\overline{G}_{\overline{a}^*|v}(2)=\overline{m}(2),a(1),m(1),l(1),v]P\{l(1)|\overline{G}_{\overline{a}^*|v}(2)=\overline{m}(2),a(1),m(1),v\} \\
& \times P\{M_{\overline{a}^*}(2)=m(2)|\overline{a}^*(1),m(1),v\}P\{m(1)|a(1),v\} \\
= & \sum_{\overline{m}(2)}\sum_{\overline{l}(1)} E[Y_{\overline{a}\overline{m}(2)\underline{G}_{\overline{a}^*|v}(3)}|\overline{G}_{\overline{a}^*|v}(2)=\overline{m}(2),\overline{a}(2),\overline{m}(2),l(1),v]P\{l(1)|a(1),m(1),v\} \\
& \times \sum_{\overline{l}^\dagger(1)} P\{M_{\overline{a}^*}(2)=m(2)|a^*(1),m(1),l^\dagger(1),v\}P(l^\dagger(1)|a^*(1),m(1),v\}P\{m(1)|a(1),v\} \\
= & \sum_{\overline{m}(2)}\sum_{\overline{l}(1)} E[Y_{\overline{a}\overline{m}(2)\underline{G}_{\overline{a}^*|v}(3)}|\overline{G}_{\overline{a}^*|v}(2)=\overline{m}(2),\overline{a}(2),\overline{m}(2),l(1),v]P\{l(1)|a(1),m(1),v\} \\
& \times \sum_{\overline{l}^\dagger(1)} P\{M_{\overline{a}^*}(2)=m(2)|\overline{a}^*(2),m(1),l^\dagger(1),v\}P(l^\dagger(1)|a^*(1),m(1),v\}P\{m(1)|a(1),v\} \\
= & \sum_{\overline{m}(2)}\sum_{\overline{l}(1)} E[Y_{\overline{a}\overline{m}(2)\underline{G}_{\overline{a}^*|v}(3)}|\overline{G}_{\overline{a}^*|v}(2)=\overline{m}(2),\overline{a}(2),\overline{m}(2),l(1),v]P\{l(1)|a(1),m(1),v\} \\
& \times \sum_{\overline{l}^\dagger(1)} P\{m(2)|\overline{a}^*(2),m(1),l^\dagger(1),v\}P(l^\dagger(1)|a^*(1),m(1),v\}P\{m(1)|a(1),v\}
\end{aligned}
$$

Note that in the expectation in the second and subsequent equalities we cannot remove $G_{\overline{a}^*|v}(1)$ from the conditioning set as it will be associated with $\underline{G}_{\overline{a}^*|v}(2)$. In the fifth inequality we can make the substition $P\{G_{\overline{a}^*|v}(2)=m(2)|G_{\overline{a}^*|v}(1)=m(1),a(1),m(1),v\}=P\{G_{\overline{a}^*|v}(2)=m(2)|G_{\overline{a}^*|v}(1)=m(1),a^*(1),m(1),v\}$ because the first expression is equal to $\frac{P\{G_{\overline{a}^*|v}(2)=m(2),G_{\overline{a}^*|v}(1)=m(1)|a(1),m(1),v\}}{P\{G_{\overline{a}^*|v}(1)=m(1)|a(1),m(1),v\}}$ and the second to $\frac{P\{G_{\overline{a}^*|v}(2)=m(2),G_{\overline{a}^*|v}(1)=m(1)|a^*(1),m(1),v\}}{P\{G_{\overline{a}^*|v}(1)=m(1)|a^*(1),m(1),v\}}$ and these latter expressions are equal to each other since $(G_{\overline{a}^*|v}(2),G_{\overline{a}^*|v}(1))$, being random

http://biostats.bepress.com/harvardbiostat/paper168

draws, will be independent of any actual observed variables. Likewise in the seventh equality, we can remove $\overline{G}_{\overline{a}^*|v}(2)$ is the conditioning set in $P\{l(1)|\overline{G}_{\overline{a}^*|v}(2) = \overline{m}(2), a(1), m(1), v\}$ because $\overline{G}_{\overline{a}^*|v}(2)$ will be independent of all actual observed variables. If we carry on with this argument iteratively we obtain:

$$
\begin{aligned}
&= \sum_{\overline{m}} \sum_{\bar{l}(T-1)} E[Y_{\overline{am}}|\overline{G}_{\overline{a}^*|v} = \overline{m}, \overline{a}, \overline{m}, \bar{l}, v] \prod_{t=1}^{T-1} P\{\bar{l}(t)|\overline{a}(t), \overline{m}(t), \bar{l}(t-1), v)\} \\
&\quad \times \sum_{\bar{l}^\dagger(T-1)} \prod_{t=1}^{T} P\{M(t)|\overline{a}(t), \overline{m}(t-1), \bar{l}^\dagger(t-1), v\} P\{\bar{l}^\dagger(t-1)|\overline{a}(t-1), \overline{m}(t-1), \bar{l}^\dagger(t-2), v)\} \\
&= \sum_{\overline{m}} \sum_{\bar{l}(T-1)} E[Y_{\overline{am}}|\overline{a}, \overline{m}, \bar{l}, v] \prod_{t=1}^{T-1} P\{\bar{l}(t)|\overline{a}(t), \overline{m}(t), \bar{l}(t-1), v)\} \\
&\quad \times \sum_{\bar{l}^\dagger(T-1)} \prod_{t=1}^{T} P\{M(t)|\overline{a}(t), \overline{m}(t-1), \bar{l}^\dagger(t-1), v\} P\{\bar{l}^\dagger(t-1)|\overline{a}(t-1), \overline{m}(t-1), \bar{l}^\dagger(t-2), v)\} \\
&= \sum_{\overline{m}} \sum_{\bar{l}(T-1)} E[Y|\overline{a}, \overline{m}, \bar{l}, v] \prod_{t=1}^{T-1} P\{\bar{l}(t)|\overline{a}(t), \overline{m}(t), \bar{l}(t-1), v)\} \\
&\quad \times \sum_{\bar{l}^\dagger(T-1)} \prod_{t=1}^{T} P\{M(t)|\overline{a}^*(t), \overline{m}(t-1), \bar{l}^\dagger(t-1), v\} P\{\bar{l}^\dagger(t-1)|\overline{a}^*(t-1), \overline{m}(t-1), \bar{l}^\dagger(t-2), v)\}.
\end{aligned}
$$

This completes the proof.

Natural Direct and Indirect Effects with a Time-Varying Exposure and Mediator but no Time-Varying Confounding

Consider the causal diagram in Figure 5 in which $A(t)$ and $M(t)$ are not time-varying and suppose this were a non-parametric structural equation model (Shpitser and Pearl, 2008; Pearl, 2009). The following assumptions would then hold: (i*) $Y_{\overline{am}} \perp\!\!\!\perp A(t)|\overline{A}(t-1), \overline{M}(t-1), V$ and (ii*) $Y_{\overline{am}} \perp\!\!\!\perp M(t)|\overline{A}(t), \overline{M}(t-1), V$ and (iii*) $M_{\overline{a}}(t) \perp\!\!\!\perp A(t)|\overline{A}(t-1), \overline{M}(t-1), V$, and (iv*) $Y_{\overline{am}} \perp\!\!\!\perp \overline{M}_{\overline{a}^*}(t)|V$. It can be shown that assumption (iv*) follows from the non-parametric structural equation model using a twin network diagram (—). Note also assumptions (i*)-(iv*) would also hold if there were a variable U_A in Figure 5 with edges into $A(t)$ for any or all t (but no edges into any $M(t)$) and/or if if there were a variable U_M in Figure 5 with edges into $M(t)$ for any or all t (but no edges into any $A(t)$). The natural direct effect can then be defined as $Y_{\overline{a}M_{\overline{a}^*}} - Y_{\overline{a}^*M_{\overline{a}^*}}$ and the natural indirect effect as $Y_{\overline{a}M_{\overline{a}}} - Y_{\overline{a}M_{\overline{a}^*}}$. We assume composition that $Y_{\overline{a}} = Y_{\overline{a}M_{\overline{a}}}$. We have the decomposition of a total effect into natural direct and indirect effects $Y_{\overline{a}} - Y_{\overline{a}^*} = (Y_{\overline{a}M_{\overline{a}}} - Y_{\overline{a}M_{\overline{a}^*}}) + (Y_{\overline{a}M_{\overline{a}^*}} - Y_{\overline{a}^*M_{\overline{a}^*}})$.

Under assumptions (i*)-(iv*), average natural direct and indirect effects conditional on $V = v$ are identified since

$$
\begin{aligned}
& E[Y_{\overline{a}M_{\overline{a}^*}}|v] \\
&= \sum_{\overline{m}} E[Y_{\overline{am}}|M_{\overline{a}^*} = \overline{m}, v] P(M_{\overline{a}^*} = \overline{m}|v) \\
&= \sum_{\overline{m}} E[Y_{\overline{am}}|v] P(M_{\overline{a}^*} = \overline{m}|v) \\
&= \sum_{\overline{m}} E[Y|\overline{a}, \overline{m}, v] \prod_{t=1}^{T} P\{M(t)|\overline{a}^*(t), \overline{m}(t-1), v\}
\end{aligned}
$$

where the final equality follows by application of Robin's g-formula (Robins, 1986). The

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

average natural direct effect conditional on $V = v$ is thus given by $E[Y_{\overline{a}M_{\overline{a}^*}}|v] - E[Y_{\overline{a}^*M_{\overline{a}^*}}|v] =$

$$\sum\nolimits_{\overline{m}} \{E[Y|\overline{a}, \overline{m}, v] - E[Y|\overline{a}^*, \overline{m}, v]\} \prod\nolimits_{t=1}^{T} P\{M(t)|\overline{a}^*(t), \overline{m}(t-1), v\}.$$

The average natural indirect effect conditional on $V = v$ is given by $E[Y_{\overline{a}M_{\overline{a}}}|v] - E[Y_{\overline{a}M_{\overline{a}^*}}|v] =$

$$\sum\nolimits_{\overline{m}} E[Y|\overline{a}, \overline{m}, v] \prod\nolimits_{t=1}^{T} [P\{M(t)|\overline{a}(t), \overline{m}(t-1), v\} - P\{M(t)|\overline{a}^*(t), \overline{m}(t-1), v\}].$$

This final expression is a generalization of Pearl's mediation formula (Pearl, 2012) for time-varying exposures and mediators.

Proposition. Consider then the following regression models:

$$\begin{aligned}
E[M(1)|m(0), y(0), \overline{a}(1), c] &= \beta_{10} + \beta_{11}a(0) + \beta_{12}a(1) + \beta_{13}m(0) + \beta_{14}y(0) + \beta'_{15}c \\
E[M(2)|\overline{m}(1), \overline{y}(1), \overline{a}(2), c] &= \beta_{20} + \beta_{21}a(1) + \beta_{22}a(2) + \beta_{23}m(1) + \beta_{24}y(1) + \beta'_{25}c \\
E[Y(1)|\overline{m}(1), y(0), \overline{a}(1), c] &= \theta_{10} + \theta_{11}a(0) + \theta_{12}a(1) + \theta_{13}m(0) + \theta_{14}m(1) + \theta_{15}y(0) + \theta'_{16}c \\
E[Y(2)|\overline{m}(2), \overline{y}(1), \overline{a}(2), c] &= \theta_{20} + \theta_{21}a(1) + \theta_{22}a(2) + \theta_{23}m(1) + \theta_{24}m(2) + \theta_{25}y(1) + \theta'_{26}c.
\end{aligned}$$

Under assumptions (i†)-(iii†) with $V = (C, A(0),\ M(0), Y(0))$ and $L(1) = Y(1)$, with two intervention periods, $A(1)$ and $A(2)$, the randomized interventional analgues of the natural direct and indirect effects are given by:

$$\begin{aligned}
E(Y_{\overline{a}G_{\overline{a}^*|v}}|v) - E(Y_{\overline{a}^*G_{\overline{a}^*|v}}|v) &= (\theta_{21} + \theta_{12}\theta_{25})[a(1) - a^*(1)] + \theta_{22}[a(2) - a^*(2)] \\
\{E(Y_{\overline{a}G_{\overline{a}|v}}|v) - E(Y_{\overline{a}G_{\overline{a}^*|v}}|v)\} &= \{\theta_{23}\beta_{12} + \theta_{25}\theta_{14}\beta_{12} + \beta_{21}\theta_{24} + \beta_{24}\theta_{12}\theta_{24}\}[a(1) - a^*(1)] \\
&\quad + \beta_{22}\theta_{24}[a(2) - a^*(2)].
\end{aligned}$$

Proof. By the mediational g-formula we have, $E[Y_{\overline{a}G_{\overline{a}^*|c}}|c] =$

$$\int_{\overline{m}} \int_{\overline{l}(T-1)} E[Y|\overline{a}, \overline{m}, \overline{l}, c] \prod\nolimits_{t=1}^{T-1} P\{\overline{l}(t)|\overline{a}(t), \overline{m}(t), \overline{l}(t-1), c)\}$$
$$\times \int_{\overline{l}^\dagger(T-1)} \prod\nolimits_{t=1}^{T} P\{M(t)|\overline{a}^*(t), \overline{m}(t-1), \overline{l}^\dagger(t-1), c\} P\{\overline{l}^\dagger(t-1)|\overline{a}^*(t-1), \overline{m}(t-1), \overline{l}^\dagger(t-2), c)\}.$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/harvardbiostat/paper168

We have that

$$
\begin{aligned}
&\int_{\bar{l}(T-1)} E[Y|\overline{a},\overline{m},\bar{l},c]\prod_{t=1}^{T-1} P\{\bar{l}(t)|\overline{a}(t),\overline{m}(t),\bar{l}(t-1),c)\} \\
= &\int_{y(1)} E[Y(2)|\overline{m}(2),\overline{y}(1),\overline{a}(2),c]P\{y(1)|\overline{m}(1),y(0),\overline{a}(1),c\} \\
= &\ \theta_{20}+\theta_{21}a(1)+\theta_{22}a(2)+\theta_{23}m(1)+\theta_{24}m(2)+\theta_{25}E[Y(1)|\overline{m}(1),y(0),\overline{a}(1),x]+\theta'_{26}c \\
= &\ \theta_{20}+\theta_{21}a(1)+\theta_{22}a(2)+\theta_{23}m(1)+\theta_{24}m(2) \\
&+\theta_{25}\{\theta_{10}+\theta_{11}a(0)+\theta_{12}a(1)+\theta_{13}m(0)+\theta_{14}m(1)+\theta_{15}y(0)+\theta'_{16}c\}+\theta'_{26}c.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\int_{\overline{m}}\int_{\bar{l}(T-1)} E[Y|\overline{a},\overline{m},\bar{l},c]\prod_{t=1}^{T-1} P\{\bar{l}(t)|\overline{a}(t),\overline{m}(t),\bar{l}(t-1),c)\} \\
&\times\int_{\bar{l}^{\dagger}(T-1)}\prod_{t=1}^{T} P\{M(t)|\overline{a}^*(t),\overline{m}(t-1),\bar{l}^{\dagger}(t-1),c\}P\{\bar{l}^{\dagger}(t-1)|\overline{a}^*(t-1),\overline{m}(t-1),\bar{l}^{\dagger}(t-2),c)\} \\
= &\ \theta_{20}+\theta_{21}a(1)+\theta_{22}a(2)+\theta_{25}\{\theta_{10}+\theta_{11}a(0)+\theta_{12}a(1)+\theta_{13}m(0)+\theta_{15}y(0)+\theta'_{16}c\}+\theta'_{26}c \\
&+\int_{\overline{m}}\{\theta_{23}m(1)+\theta_{24}m(2)+\theta_{25}\theta_{14}m(1)\} \\
&\times\int_{\bar{l}^{\dagger}(T-1)}\prod_{t=1}^{T} P\{M(t)|\overline{a}^*(t),\overline{m}(t-1),\bar{l}^{\dagger}(t-1),c\}P\{\bar{l}^{\dagger}(t-1)|\overline{a}^*(t-1),\overline{m}(t-1),\bar{l}^{\dagger}(t-2),c)\}.
\end{aligned}
$$

$$
\begin{aligned}
= &\ \theta_{20}+\theta_{21}a(1)+\theta_{22}a(2)+\theta_{25}\{\theta_{10}+\theta_{11}a(0)+\theta_{12}a(1)+\theta_{13}m(0)+\theta_{15}y(0)+\theta'_{16}c\}+\theta'_{26}c \\
&+\{\theta_{23}+\theta_{25}\theta_{14}\}E[M(1)|m(0),y(0),\overline{a}^*(1),c] \\
&+\theta_{24}\int_{y(1)} E[M(2)|\overline{m}(1),\overline{y}(1),\overline{a}^*(2),c]P\{\overline{y}(1)|\overline{m}(1),y(0),\overline{a}^*(1),c\} \\
= &\ \theta_{20}+\theta_{21}a(1)+\theta_{22}a(2)+\theta_{25}\{\theta_{10}+\theta_{11}a(0)+\theta_{12}a(1)+\theta_{13}m(0)+\theta_{15}y(0)+\theta'_{16}c\}+\theta'_{26}c \\
&+\{\theta_{23}+\theta_{25}\theta_{14}\}\{\beta_{10}+\beta_{11}a(0)+\beta_{12}a^*(1)+\beta_{13}m(0)+\beta_{14}y(0)+\beta'_{15}c\} \\
&+\theta_{24}\int_{y(1)}\{\beta_{20}+\beta_{21}a^*(1)+\beta_{22}a^*(2)+\beta_{23}m(1)+\beta_{24}y(1)+\beta'_{25}c\}P\{\overline{y}(1)|\overline{m}(1),y(0),\overline{a}^*(1),c\} \\
= &\ \theta_{20}+\theta_{21}a(1)+\theta_{22}a(2)+\theta_{25}\{\theta_{10}+\theta_{11}a(0)+\theta_{12}a(1)+\theta_{13}m(0)+\theta_{15}y(0)+\theta'_{16}c\}+\theta'_{26}c \\
&+\{\theta_{23}+\theta_{25}\theta_{14}\}\{\beta_{10}+\beta_{11}a(0)+\beta_{12}a^*(1)+\beta_{13}m(0)+\beta_{14}y(0)+\beta'_{15}c\} \\
&+\theta_{24}\{\beta_{20}+\beta_{21}a^*(1)+\beta_{22}a^*(2)+\beta_{23}m(1)+\beta_{24}E[Y(1)|\overline{m}(1),y(0),\overline{a}^*(1),c]+\beta'_{25}c\} \\
= &\ \theta_{20}+\theta_{21}a(1)+\theta_{22}a(2)+\theta_{25}\{\theta_{10}+\theta_{11}a(0)+\theta_{12}a(1)+\theta_{13}m(0)+\theta_{15}y(0)+\theta'_{16}c\}+\theta'_{26}c \\
&+\{\theta_{23}+\theta_{25}\theta_{14}\}\{\beta_{10}+\beta_{11}a(0)+\beta_{12}a^*(1)+\beta_{13}m(0)+\beta_{14}y(0)+\beta'_{15}c\} \\
&+\theta_{24}[\beta_{20}+\beta_{21}a^*(1)+\beta_{22}a^*(2)+\beta_{23}m(1) \\
&\qquad +\beta_{24}\{\theta_{10}+\theta_{11}a(0)+\theta_{12}a^*(1)+\theta_{13}m(0)+\theta_{14}m(1)+\theta_{15}y(0)+\theta'_{16}c\}+\beta'_{25}c]
\end{aligned}
$$

Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Thus the randomized interventional analogue of the natural direct effect is given by:

$$E(Y_{\bar{a}G_{\bar{a}^*|v}}|v) - E(Y_{\bar{a}^*G_{\bar{a}^*|v}}|v) = (\theta_{21} + \theta_{12}\theta_{25})[a(1) - a^*(1)] + \theta_{22}[a(2) - a^*(2)]$$

and the randomized interventional analogue of the natural direct effect is given by:

$$\begin{aligned}\{E(Y_{\bar{a}G_{\bar{a}|v}}|v) - E(Y_{\bar{a}G_{\bar{a}^*|v}}|v)\} &= \{\theta_{23}\beta_{12} + \theta_{25}\theta_{14}\beta_{12} + \beta_{21}\theta_{24} + \beta_{24}\theta_{12}\theta_{24}\}[a(1) - a^*(1)] \\ &\quad +\beta_{22}\theta_{24}[a(2) - a^*(2)].\end{aligned}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/harvardbiostat/paper168